6-2014

Socio-physical analytics: Challenges & opportunities

Archan MISRA
Singapore Management University, archanm@smu.edu.sg

Kasthuri JAYARAJAH
Singapore Management University, kasthurij.2014@phdis.smu.edu.sg

Shriguru NAYAK
Singapore Management University, shrigurun@smu.edu.sg

Philips Kokoh PRASETYO
Singapore Management University, pprasetyo@smu.edu.sg

Ee-peng LIM
Singapore Management University, eplim@smu.edu.sg

DOI: https://doi.org/10.1145/2611264.2611265

Citation

MISRA, Archan; JAYARAJAH, Kasthuri; NAYAK, Shriguru; PRASETYO, Philips Kokoh; and Ee-peng LIM. Socio-physical analytics: Challenges & opportunities. (2014). *WPA '14: Proceedings of the 2014 Workshop on Physical Analytics, Bretton Woods, New Hampshire, June 16*. 19-24. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/3637

Socio-Physical Analytics: Challenges & Opportunities

Archan Misra, Kasthuri Jayarajah, Shriguru Nayak, Philips Kokoh Prasetyo and Ee-Peng Lim
Singapore Management University

ABSTRACT

In this paper, we argue for expanded research into an area called *Socio-Physical Analytics*, that focuses on combining the behavioral insight gained from mobile-sensing based monitoring of physical behavior with the inter-personal relationships and preferences deduced from online social networks. We highlight some of the research challenges in combining these heterogeneous data sources and then describe some examples of our ongoing work (based on real-world data being collected at SMU) that illustrate two aspects of socio-physical analytics: (a) how additional demographic and online analytics based attributes can potentially provide better insights into the preferences and behaviors of individuals or groups (in terms of movement prediction and understanding of physical vs. online interactions), and (b) how online and physical interactions can help us discover latent characteristics of physical spaces and entities.

Categories and Subject Descriptors

H.3.4 [**Systems and Software**]: Information Networks

Keywords

mobile sensing; information fusion; social media analytics

1. INTRODUCTION

Mobile sensing research over the past decade has principally focused on accurately recognizing the *physical* characteristics of user activities, such as *where* an individual is (exemplified by the active research on indoor location tracking), and *what physical activity* she is performing (exemplified by research on low and high-level activity recognition). In our view, especially in dense urban centers such as Singapore, where users spend a significant part of their daily lives in indoor public spaces (such as malls, subway stations, airports and university campuses), such location/movement and activity traces embody something more fundamental: *they carry latent insights about the semantics of activities and preferences of the individuals, their relationships with one another and the space they inhabit*. For example, instead of just inferring that a "group of 5 people are sitting in a campus coffee shop", the sensing data can conceivably be used to infer deeper *semantic* contexts–such as *why* they were sitting (e.g., to have a quick bite to eat, to discuss some class project or to plan for an upcoming musical event), and *what/when* they were likely to do next (e.g., go to class together in 5 mins, head to the gym separately in 20 mins, etc.).

Our key contention is that such deeper analytics-driven understanding will, however, only be possible if the captured traces of such mobility and sensor data can be judiciously *fused* with the dramatic recent advances in social media sensing & analytics. However, such fusion is not simply a matter of getting access, or combining at an *individual-level*, multiple mobile and online-sensing based data; a new class of analytics tools is needed to (a) handle the diverse characteristics of data generated from mobile and online sensing, as well as the differences in human interaction between the online and the physical worlds, and (b) support more scalable, real-time processing when fusing tens of thousands of such disparate mobile and online data streams.

Accordingly, in this paper, we articulate a research agenda for a field we call **Socio-Physical Analytics**, which focuses on real-time extraction of deeper individual and group-level insights from a *combination* of (a) mobile-generated 'sensor' data (capturing physical activity characteristics), (b) varying levels of demographic attributes (either captured directly or inferred via social media sensing) and (c) social media-generated 'online' data. Our research efforts are specially oriented towards densely-occupied, indoor public spaces, which possess a set of distinct characteristics that makes it different from prior research focusing on city-scale insights (e.g., [1]) generated from outdoor movement data traces. An important goal of ours is to direct the research communities' attention towards efforts for better *information fusion and discovery* from the existing diverse data sources–in many practical cases of interest, the power of longitudinal and multi-modal observations can adequately compensate for existing inaccuracies in mobile sensing (e.g., errors in indoor localization).

To articulate this research agenda, we shall first discuss some of the general characteristics of the "Socio-Physical Analytics" problem space, identifying key research challenges, especially as they relate to the ability to perform such analytics processing in near-real time over public space-scale (i.e., tens of thousands of sensor streams) data inputs. We shall then provide early examples of our work in extracting semantic insights from the combined fusion of mobile, online and demographic information sources, and show how the key research challenges are manifested in our work.

2. RELATED WORK

As this paper is intended to articulate a broad research agenda, we briefly summarize only some of the key trends on mobile and

WPA'14, June 16, 2014, Bretton Woods, New Hampshire, USA.

http://dx.doi.org/10.1145/2611264.2611265.

online analytics research that relate to our empirical problems discussed later.

Mobile Sensing & Analytics: Mobile sensing and analytics has been characterized by a progressive migration from sensing (e.g., Jigsaw [2]) of low-level activities (such as walking or cycling), to progressively higher semantic activities (such as shopping or queuing). While articulating a similar vision of combining social media and mobile sensing data, the *SocialFusion* system [3] focuses mostly on deriving more detailed *individual-level* context information, and not so much on capturing or understanding interrelated human behavior.

Movement Models and Characterization: Approaches such as [4] have taken into account location semantics (such as home vs. office) and social relationships for outdoor movement prediction. Indoor movement models have been largely statistical (e.g., [5]), and do not take into account deeper social ties among multiple participants. Recently suggested location prediction models (e.g., [6]) exploit correlations over the movement of other individuals, but without considering the *type* and multi-modal nature of inter-personal relationships.

Online Relationship Mining: The recent research on online social networks has focused on problems such as structural characterization (e.g., [7]) or distinguishing weak ties from strong ties [8]. While social ties can be observed online, the combined use of social media and physical world interactions to infer deeper relationships (especially using the fine-grained behavioral insight captured through physical sensing) has not been extensively addressed. For example, the arrival pattern at a restaurant may provide fine-grained insight into the relationship asymmetry between two individuals.

3. SMU DATASETS FOR PHYSICAL & ONLINE ANALYTICS

Before detailing the research agenda, we briefly describe the experimental testbed and analytics systems that provide preliminary data for our early research experiences.

The location and demographic data related to the behavior of individuals in indoor public spaces is collected using SMU's LiveLabs testbed [9]. LiveLabs is a large-scale testbed effort, with deployments planned in several public spaces, with the goal of collecting and inferring deep-context in near-real time from thousands of users, to enable real world experimentation with context-aware mobile applications and services. As part of this operation, LiveLabs has collected indoor location traces (using server-side RF measurements) of *all* mobile devices (from over 8000 students and 1000 staff members) that attach to the campus Wi-Fi network at SMU since September, 2013; the location readings are obtained every 2-3 minutes with approx. 8 - 10 m accuracy. Moreover, about 800 out of 2000+ opted-in LiveLabs participants have provided an extended set of demographic information, such as the school, year of study, and gender–the results presented here involve data from 828 students (450 males, 378 females) from all of SMU's six schools and spanning all 4 years of undergraduate study.

A parallel research effort under the joint SMU-CMU Living Analytics Research Center (LARC) has developed the *Palanteer* system [10]: a suite of real-time social media search and analytics tools that gathers Twitter and Foursquare feed data generated by about 150K and 200K active users in Singapore. The analytics capabilities of these tools include a less-than-five minutes delay in indexing tweets and check-in's, classification of online and off-line friendships, automated extraction of key-phrases from tweet content and detection of emerging events reported by the user community. Additionally, the researchers have developed an efficient probabilistic *network linkage* tool, which utilizes common patterns of interactions in the Twitter and Foursquares network to identify the common individuals (but with distinct IDs) across these networks. Figure 1 shows a visualization of the so-called "ego-networks" of a matched Twitter-Foursquare user pair.

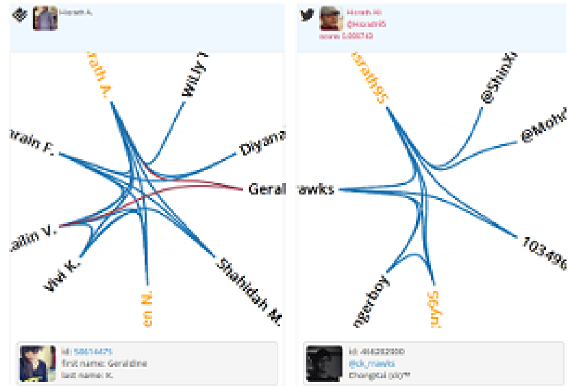

Figure 1: Network linkage (across Twitter & Foursquare)

4. RESEARCH CHALLENGES

We hypothesize that the combination of mobile sensing-based physical context and social media-based online context will provide capabilities for far richer and deeper understanding of human activities and preferences. As a single, easy-to-understand illustration of our socio-physical analytics vision, consider this example:

A well-known pop music star stops by for a surprise campus visit on a Thursday. An analysis of the resulting tweets (by the Palanteer system) generated in the vicinity of the campus can help us discover this unexpected event, as well as additional metadata, such as the (start,end) times and the on-campus location of the star's visit. On the other hand, analysis of the physical movement traces of the campus inhabitants can reveal, for example, that person A was a member of an ad-hoc group (consisting of A and three other group members who are part of the university's pop music club) that remained stationary (in the same campus location that the pop star visited) during the bulk of this visit period. Moreover, historical data does not show this group as spending time together on past Thursdays at that time.

From these facts, we can, with strong likelihood, infer that individual A and her friends were together that Thursday to follow the pop star–revealing not just individual-level activities, but also illuminating their inter-personal relationships.

4.1 Key Research Challenges

Combining the analytics insight from the mobile-sensing based world of physical activities and the social media-based world of online content generation and distribution requires us to address several key challenges.

Multi-Modal Networks and Heterogeneity: Users interact and build social ties in both physical and online-worlds. Other than user-user interactions, users also interact with other entities (sometimes known as *foci*), such as having meals at restaurants, attending schools and shopping at retail stores. Users, entities and their multiple forms of relationships and interactions can be modeled as multi-modal networks, instead of the familiar social network representation that only captures social ties among nodes of the same kind. Researchers have recently begun to study *social-affiliation* *networks*, a specialized type of multi-modal networks consisting of both user and entity nodes with social ties between users and adoption ties between users and entities. However, such multi-modal networks become even more complex when we consider interactions between users in the physical world.

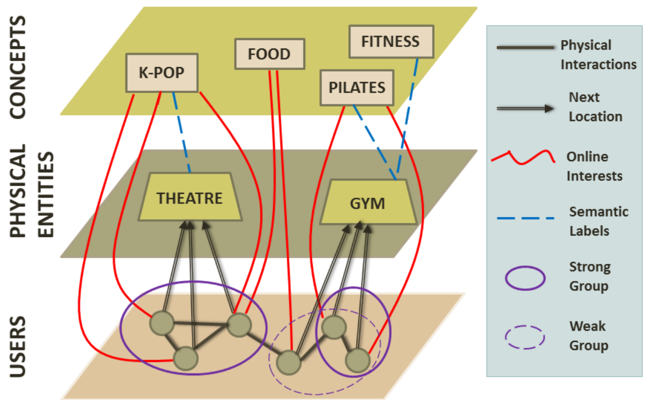

Figure 2: A multi-modal network representation of socio-physical groups

There are several sub-challenges associated with these multi-modal networks. Firstly, there is much room to empirically study and discover new properties for these networks (small world, power law degree distribution and other well known properties of uni-modal social networks may not always hold in this new network genre), and utilize such properties for better discovery of interactions among users. Related to the above challenge at the network level is the research challenge of studying behaviors at the user level in multi-modal networks, as users can have multiple personas in the physical and online worlds. Such personas can also be time-varying. For example, a student may spend much more time browsing his online course materials than visiting pubs during the examination period. This behavior pattern may completely reverse after the examination. Learning a user's behavior patterns (and their inter-relationships) in multi-modal networks is therefore an interesting challenge.

Moreover, when defining links between online entities and physical spaces in multi-modal networks, research is needed to define and utilize appropriate associations among these interaction relationships. Examples of such associations will need to capture both *complementary* (e.g., searching for restaurants on the mobile Web and dining at a restaurant in the mall) and *competing* (e.g., playing games online while attending a physical lecture on campus) interactions–something that has not been considered till date.

Dynamic and Multi-Time Scale Relationships: In the online world, social connections between individuals and user interests are typically *stable* and evolve over different time-scales. For example, while the list of friends in one's mobile phone contact list certainly evolves, it does so on time scales of days or weeks. There are, of course, exceptions: our studies do show that some Twitter users adopt new hashtags and de-adopt them within 24 hours [11], especially related to large live events. Accordingly, slow evolving user properties (such as the number of friends, or cliques in a contact graph, and preferences) are often pre-computed offline and updated with long time intervals. Relationships in the physical world and event oriented user interests can be much more dynamic, evolving at much smaller time-scales. For example, an individual can be part of both certain repeating groups (e.g, a study group that meets twice every week during a semester) that have long-term stability, as well as other more dynamic and transient groups (e.g., a group of volunteers who interact extensively during a 2-day event). Clearly, in such cases, properties of a contact graph need to be computed (or updated) more dynamically. Note that even slow-timescale properties in the social world may impact the insights obtained from faster-timescale physical world behavior. As an example, Figure 2 illustrates the problem of group detection: while analysis of traveling companions in the physical world may provide initial candidates for dynamic groups, the inclusion of online relationship-based linkages may significantly increase the *confidence* of such group detection.

A central concept in inter-personal relationships is the notion of the *strength of ties*. In typical online social media-based networks, tie strength is typically based on the intensity of interactions observed over a longer period (e.g., a week or a month). On the other hand, for the dynamic physical groups, strong ties may hold for relatively short durations of intense interactions between two users (e.g., during a 2-day event they're co-organizing), whereas weaker ties may remain stable over longer durations. A particularly interesting open problem relates to ways to define multi-temporal scale metrics for expressing such strength of ties. Note that, our very preliminary analysis of results (discussed later in Section 5.2) indicates that online and physical-world interactions *may* be distinct. Significant new research is thus needed to determine the best way to combine effects from tie-strengths from multiple physical and online social media-based interactions. For example, to alert an individual about some new content (either online content or information about some physical event on the campus), should we lay more emphasis on the consumption patterns/interests of the more stable "friends" in the online world, or the more-fleeting-but-intense "peers" in the physical world?

Real-time Multi-modal Analytics: The rise of socio-physical analytics also often accelerates the latency requirements associated with the analytics processing. Currently, most of the social media-based analytics infrastructures (including the Palanteer system at LARC) employ a batched, quasi-offline processing paradigm (utilizing processing platforms such as Hadoop). In contrast, most mobile sensing analytics for activity recognition is performed on a per-individual basis in "real-time" (with latencies on the order of tens of seconds), utilizing large-scale stream processing platforms (such as IBM InfoSphere Streams).

For socio-physical analytics, however, the processing is no longer confined to data streams of a single type (or connected by a uni-modal graph relationship). There exists substantial evidence (e.g., [12]) that implementing stream processing algorithms for low-latency processing of multi-modal data streams is a non-trivial process, requiring careful design of the data processing pipeline and involving tradeoffs between the accuracy/comprehensiveness of the processing of results and the input load. For many examples of physical analytics (such as providing more intelligent context-aware recommendations to LiveLabs participants in shopping malls and the airport), the processing latency has to remain relatively low (e.g., 10-15 secs), while handling non-negligible numbers of data streams (the SMU campus can have location data from over 12,000 devices simultaneously, whereas our target airport has over 120,000 visitors on a typical day).

5. EARLY EXAMPLES OF ANALYTICS

We now describe 3 distinct analytics problems that we are currently addressing. These problems have a common theme of augmenting mobile sensing-based physical location/movement data with additional demographic or semantic label data to derive deeper insights into the behavior of individuals in public spaces.

5.1 Movement Models and Prediction

In this research problem, our goal is to try and build better indoor mobility models for individuals, and use this to predict features such as the likely residency time in the current location and the next location to be visited. Clearly, such predictions are important for a whole variety of context-enabled applications (e.g., recommendation systems for stores and restaurants in shopping malls). Existing models of indoor movement do not explore in detail if such movement is affected by other online or demographic attributes–for example, by the relationships between a group of people moving together. To explore such possibilities, we have studied the movement data collected on the LiveLabs@SMU testbed for the influence of two additional attributes–(i) the size of the current group to which an individual belongs, and (ii) the year of study and/or gender of the individual. In particular, we detected traveling companions [13] from daily location traces (using a technique that was externally validated using movement traces of 154 users) and observed (1) the distribution of stay times at "places", and (2) the distribution of "next places" visited as functions of the group size.

Movement Behavior: Figure 3 shows the CDF of the number of popular locations visited by individuals and groups on one of the busiest days on campus. We see that nearly 70% of all individuals visited more than three distinct on-campus locations, whereas most groups sized 2, 3 and 4 visited at most three locations. We also see that larger groups (with more than 4 members) visited at most two locations. For our campus environment, our intuition is that individuals exhibit a propensity to be resident at a specific location for a shorter period of time, whereas larger groups tend to spend longer times at fewer number of locations. Of course, this behavior may be venue dependent–e.g., airports may have highly-mobile large tour groups.

Next Place Prediction: We also studied the transition probabilities of individuals vs. groups of different sizes (2, 3 and 4) in terms of the semantics of the next place they visit (such as a class room, a meeting room, a common area) after leaving the central food court on the campus. Table 1 shows the results of the paired t-test between these transition probability distributions for individuals vs. different-sized groups, revealing that there is a clear statistical difference between the transition behavior of individuals vs. groups, with the difference becoming more significant with larger group sizes. Similar studies show that gender, on the other hand, does not play as significant a discriminatory role (t-test values of males vs. females was $\tilde{0}.43$), whereas the 'year of study' turned out to be significant in the case of 3^{rd} vs. 4^{th} year students. While our results are essentially statistical (and do not provide any evidence that such attributes *cause* the resulting differences in transition probabilities), these findings show that individual-level attributes may not always be sufficient for understanding physical-world human behavior.

	Individuals	2-members	3-members	4-members
Individuals	-	8.28E-04	2.37E-08	3.95E-13
2-members	5.70E-04	-	9.43E-06	6.72E-09
3-members	4.45E-08	9.41E-06	-	3.70E-02
4-members	1.03E-13	9.53E-09	3.66E-02	-

Table 1: Student's paired t-test probabilities between distributions of "next places" from food court

While preliminary, these results suggest that it is important to develop more "demographics and social relationship"-aware movement models for better prediction of indoor movement behavior; accordingly, this is an active part of our own socio-physical analytics research agenda.

5.2 Online vs. Offline "Friendships"

The problem of distinguishing between online relationships (traditionally derived through mining of social network interaction data) and physical world (offline) relationships is one of deep interest to social scientists and behavioral experimenters. We have thus been conducting investigations into whether an "online friend is an on-campus friend as well?", or not. To study such properties, albeit in a *preliminary and inconclusive fashion*, we created two types of implied social relationships: (i) We extracted a dataset of 32 LiveLabs@SMU participants who provide "phone contact list" information, resulting in 112 pairs of asymmetric *"contact pairs"*. (ii) Separately, we had information from 48 students who were referred by other LiveLabs participants through a "referral program"; we call such pairs as *"friend pairs"*, based on the assumption that referrals implied friendship.

Figures 4 and 5 show the CDFs for amount of time for which the "contact pairs" and "friend pairs" were detected together as traveling companions (indicating that they were interacting in the physical world as well), respectively, on one of the busiest days (8th January, 2014). The figures also show the amount of time the students spent alone, and the amount of time they spent with students who were neither their contacts, nor their friends. Here, the x-axis corresponds to the time spent alone or with a contact/friend, or with others, as a percentage of the total time for which a student was spotted on that day. From Figure 4, we see that less than 20% students in our sample spent more than 70% of their time on-campus unaccompanied. Also, about 20% of the students spent more than 80% of their time in groups with students who were not on their contact lists. Surprisingly, almost all of the students spent less than 5% of their time on campus with their contacts (who are LiveLabs participants as well). On the contrary, in Figure 5, we see that at least 10% of the students spent more than 10% of their time with their friends. In fact, at least 5% of the students spent between 35 - 85% of their time with their friends.

Note that, in either case, a significant fraction of the time was spent with "others" (as our observations were limited to only those who received referrals or the contacts who are themselves LiveLabs participants, it is certainly possible (and in fact *likely*) that many of those other individuals could also be friends or 'contacts'). While our results are not quantitatively conclusive, they do suggest that there *may* be differences between 'online ties' (at least as represented through contact lists) and 'offline ties' (represented by the amount of time spent collocated on the campus). Using advanced versions of the network linkage tool (see Section 3), it may become possible to link identities across the online and physical worlds revealing more ground truth links, thus obviating the need for complete contact lists. Further, when the social media identities of the LiveLabs participants become available in future, the study can be extended to utilize the richer "online ties" as well. As mentioned before, the temporal dynamics of the relationships (did two people spend 1 hour at a stretch together, or did they meet 6 times for 10 minutes each?) may be another important factor in modeling their 'strength of ties'. *Developing better multi-modal (and multi-timescale) models of relationships between individuals is thus a key goal of our own research.*

5.3 Semantic Labeling of Spaces

In this research thread, we seek to obtain deeper and organic understanding of the characteristics of various places, by discov-

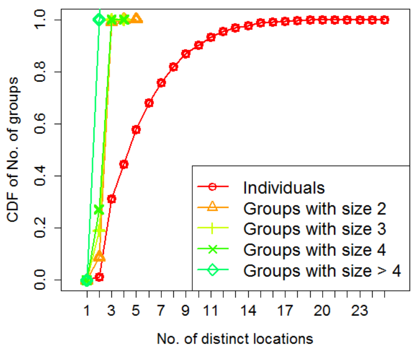

Figure 3: CDF of no. of popular locations

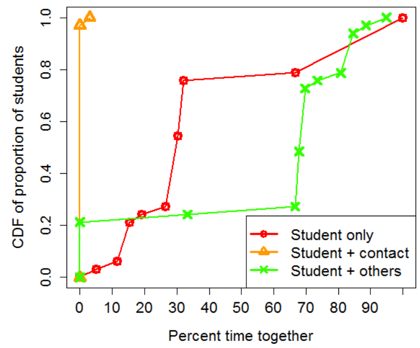

Figure 4: CDF of time spent together between contact pairs

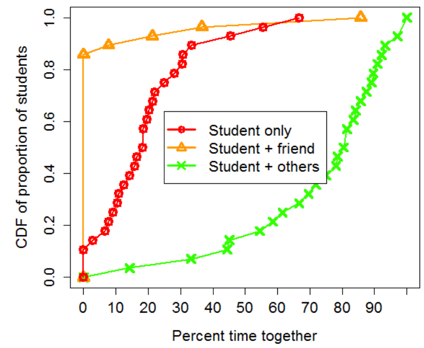

Figure 5: CDF of time spent together between referral (friend) pairs

ering their "semantic labels" (such as different informal activities conducted at various campus locations). As a first step, LARC researchers have used geo-coded tweets to determine characteristics of various places in Singapore. More specifically, the geo-coded tweets were mined to discover latent topics, which were then associated with corresponding places located within a 200m radius of the tweet's location, thereby enabling us to discover user sentiments and activities related to specific establishments (such as stores and schools). Going forward, we are now working to take additional advantage of physical movement patterns and behavior, as extracted from mobile sensing data.

Very specifically, in the SMU campus, we are additionally utilizing *residency/movement* patterns of groups and individuals to uncover the likely informal activities conducted at various on-campus locations. Our core insight is to look for similar *socio-temporal* patterns of visits (i.e., visits by the same group of individuals at around the same times on different days) to different places (some of which have a-priori activity labels formally assigned, while others don't), to discover likely similarities between activities conducted at different place pairs. As a simple example, if users A, B and C visit the *Cafeteria* as a group (with a formal label of "lunch") four days a week between 12-12:30pm, and also visit the *Library* on the other weekday from 12:15-12:40pm, one may suspect that the group is having lunch in the library (while perhaps studying together), provided that the group members did not also visit the cafeteria around that time. A large set of similar observations might then lead us to associate the informal activity of "lunch" with the Library.

Formally, this approach may be codified through a graph-based representation (see Figure 6) where the vertices denote the different on-campus locations, each with a set of formal *Activity* labels. The edges are then formed and weighed based on socio-temporal similarity metrics (e.g., a group G1 may associate a directed label with weight 0.2 between *Cafeteria* and *Library* if the group is observed to visit the *Library* with an average frequency that is 1/4th of their Cafeteria visits at similar times). Finally, novel *weighted label propagation* algorithms are applied over this multi-modal physical activity-based graph (with different groups assigning different weights and labels to the corresponding edge) to probabilistically infer the evolving set of informal activities at different locations. While our results are preliminary but encouraging, our research illustrates the necessity, of *developing an ingenious mix of online (in this case, the label propagation technique) and physical analytics techniques, to derive deeper insights from such multi-modal data*.

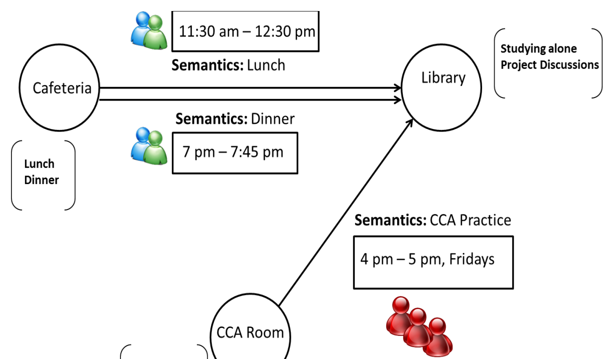

Figure 6: Discovery of uses of spaces

6. CONCLUSION

We have articulated a vision of *Socio-Physical Analytics*, which seeks to extract deeper semantic insights about people's activities and use of public spaces by fusing physical activity analytics (via mobile sensing) with demographic data and online analytics (via social media sensing). We presented 3 ongoing research activities that illustrate how such combined sensing can foster better physical activity models (specifically, movement prediction), enable better understanding of inter-personal ties (specifically, online vs. physical world interaction) and support better understanding of informal space usage, situating these specific activities in the context of a broader set of research challenges. A key take-away here is the need for deeper interaction and inter-disciplinary research spanning the mobile sensing and online analytics communities.

7. ACKNOWLEDGEMENTS

This material is based on research sponsored in part by the Air Force Research Laboratory, under agreement number FA2386-14-1-0002, and partially supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office, Media Development Authority (MDA). The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government.

8. REFERENCES

[1] Richard Becker, Ramón Cáceres, Karrie Hanson, Sibren Isaacman, Ji Meng Loh, Margaret Martonosi, James Rowland, Simon Urbanek, Alexander Varshavsky, and Chris Volinsky. Human mobility characterization from cellular network data. *Commun. ACM*, 56(1), Jan 2013.

[2] Hong Lu, Jun Yang, Zhigang Liu, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. The jigsaw continuous sensing engine for mobile phone applications. In *ACM SenSys'10*, 2010.

[3] Aaron Beach, Mike Gartrell, Xinyu Xing, Richard Han, Qin Lv, Shivakant Mishra, and Karim Seada. Fusing mobile, sensor, and social data to fully enable context-aware computing. In *ACM HotMobile*, 2010.

[4] Chiara Boldrini and Andrea Passarella. Hcmm: Modelling spatial and temporal properties of human mobility driven by users' social relationships. *Computer Communications*, 33(9):1056 – 1074, June 2010.

[5] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, Seong Joon Kim, and Song Chong. On the levy-walk nature of human mobility. *IEEE/ACM Trans. Netw.*, 2011.

[6] Manlio De Domenico, Antonio Lima, and Mirco Musolesi. Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing*, 9(6):798 – 807, 2013.

[7] Luca Maria Aiello, Alain Barrat, Rossano Schifanella, Ciro Cattuto, Benjamin Markines, and Filippo Menczer. Friendship prediction and homophily in social media. *ACM Transactions on Web*, 6(2):9:1–9:33, June 2012.

[8] Jure Leskovec and Eric Horvitz. Planetary-scale views on a large instant-messaging network. In *In WWW '08*.

[9] Rajesh Krishna Balan, Archan Misra, and Youngki Lee. Livelabs: Building an in-situ real-time mobile experimentation testbed. *ACM HotMobile*, 2014.

[10] Ee-Peng Lim and Palakorn Achananuparp. Palanteer: A search engine for community generated microblogging data. In *14th International Conference on Asia-Pacific Digital Libraries*, pages 239–248. 2012.

[11] Su Mon Kywe, Tuan-Anh Hoang, Ee-Peng Lim, and Feida Zhu. On recommending hashtags in twitter networks. In *International Conference on Social Informatics*, 2012.

[12] Kun-Lung Wu, Kirsten W. Hildrum, Wei Fan, Philip S. Yu, Charu C. Aggarwal, David A. George, Buğra Gedik, Eric Bouillet, Xiaohui Gu, Gang Luo, and Haixun Wang. Challenges and experience in prototyping a multi-modal stream analytic and monitoring application on system s. In *In VLDB '07*.

[13] Lu-An Tang, Yu Zheng, Jing Yuan, Jiawei Han, Alice Leung, Chih-Chieh Hung, and Wen-Chih Peng. On discovery of traveling companions from streaming trajectories. In *In ICDE '12*.